BE

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/62, 15/63, 15/70, C12P 21/00, 21/04, C07K 19/00, 1/113, 1/14, 1/22

A1

(11) International Publication Number: **WO 00/47751**

(43) International Publication Date: 17 August 2000 (17.08.00)

(21) International Application Number: PCT/US00/02764

(22) International Filing Date: 2 February 2000 (02.02.00)

(30) Priority Data:
09/249,543 12 February 1999 (12.02.99) US

(71) Applicant: NEW ENGLAND BIOLABS, INC. [US/US]; 32 Tozer Road, Beverly, MA 01915 (US).

(71)(72) Applicants and Inventors: EVANS, Thomas, C. [US/US]; 1st floor, 68 Albion Street, Somerville, MA 02143 (US). XU, Ming-Qun [CN/US]; 40 Crescent Road, Hamilton, MA 01982 (US).

(74) Agent: WILLIAMS, Gregory, D.; New England Biolabs, Inc., 32 Tozer Road, Beverly, MA 01915 (US).

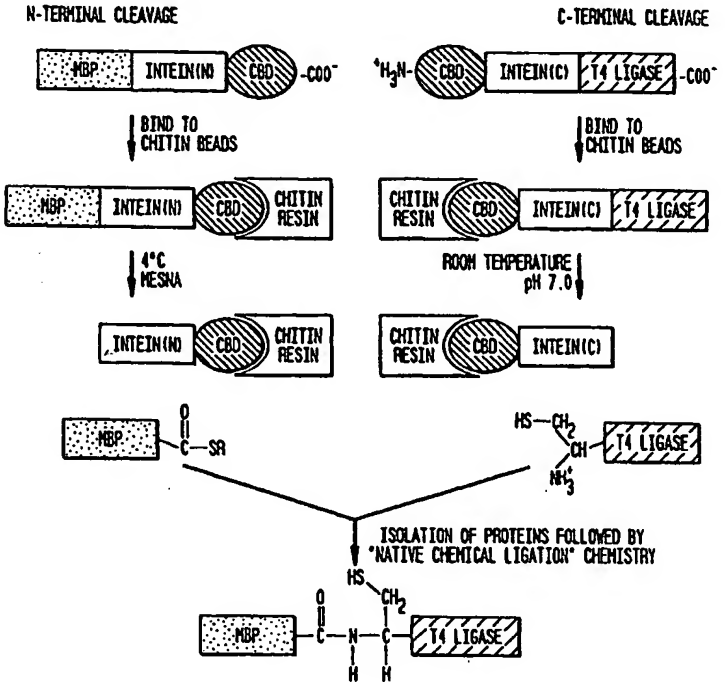
(81) Designated States: JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: INTEIN-MEDIATED PROTEIN LIGATION OF EXPRESSED PROTEINS

(57) Abstract

A method for the ligation of expressed proteins which utilizes inteins, for example the RIR1 intein from *Methanobacterium thermotrophicum*, is provided. Constructs of the *Mth* RIR1 intein in which either the C-terminal asparagine or N-terminal cysteine of the intein are replaced with alanine enable the facile isolation of a protein with a specified N-terminal, for example, cysteine for use in the fusion of two or more expressed proteins. The method involves the steps of generating a C-terminal thioester-tagged target protein and a second target protein having a specified N-terminal via inteins, such as the modified *Mth* RIR1 intein, and ligating these proteins. A similar method for producing a cyclic or polymerized protein is provided. Modified inteins engineered to cleave at their C-terminus or N-terminus, respectively, and DNA and plasmids encoding these modified inteins are also provided.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTEIN-MEDIATED PROTEIN LIGATION OF EXPRESSED PROTEINS

RELATED APPLICATIONS

This Application is a Continuation-In-Part of U.S.S.N. 08/811,492, filed March 5, 1997 now U.S. Patent No. 5,834,247, issued November 10, 1998, entitled "Modified Proteins Comprising Controllable Intervening Protein Sequences Or Their Elements Methods of Producing Same and Methods For Purification Of A Target Protein Comprised By A Modified Protein", and of U.S.S.N. 60/102,413, filed September 30, 1998, entitled "Intein Mediated Peptide Ligation."

BACKGROUND OF THE INVENTION

The present invention relates to methods of intein-mediated ligation of proteins. More specifically, the present invention relates to intein-mediated ligation of expressed proteins containing a predetermined N-terminal residue and/or a C-terminal thioester generated via use of one or more naturally occurring or modified inteins. Preferably, the predetermined residue is cysteine.

Inteins are the protein equivalent of the self-splicing RNA introns (see Perler et al., *Nucleic Acids Res.* 22:1125-1127 (1994)), which catalyze their own excision from a precursor protein with the concomitant fusion of the flanking protein sequences, known as exteins (reviewed in Perler et al., *Curr.*

Opin. Chem. Biol. 1:292-299 (1997); Perler, F. B. *Cell* 92(1):1-4 (1998); Xu et al.,*EMBO J.* 15(19):5146-5153 (1996)).

Studies into the mechanism of intein splicing led to the development of a protein purification system that utilized thiol-induced cleavage of the peptide bond at the N-terminus of the *Sce* VMA intein (Chong et al., *Gene* 192(2):271-281 (1997)). Purification with this intein-mediated system generates a bacterially-expressed protein with a C-terminal thioester (Chong et al., (1997)). In one application, where it is described to isolate a cytotoxic protein, the bacterially expressed protein with the C-terminal thioester is then fused to a chemically-synthesized peptide with an N-terminal cysteine using the chemistry described for "native chemical ligation" (Evans et al., *Protein Sci.* 7:2256-2264 (1998); Muir et al., *Proc. Natl. Acad. Sci. USA* 95:6705-6710 (1998)).

This technique, referred to as "intein-mediated protein ligation" (IPL), represents an important advance in protein semi-synthetic techniques. However, because chemically-synthesized peptides of larger than about 100 residues are difficult to obtain, the general application of IPL is limited by the requirement of a chemically-synthesized peptide as a ligation partner.

IPL technology would be significantly expanded if an expressed protein with a predetermined N-terminus, such as cysteine, could be generated. This would allow the fusion of one

or more expressed proteins from a host cell, such as bacterial, yeast or mammalian cells.

One method of generating an N-terminal cysteine is with the use of proteases. However, proteases have many disadvantages, such as the possibility of multiple protease sites within a protein, as well as the chance of non-specific degradation. Furthermore, following proteolysis, the proteases must be inactivated or purified away from the protein of interest before proceeding with IPL. (Xu, et al., *Proc. Natl. Acad. Sci. USA* 96(2):388-393 (1999) and Erlandson, et al., *Chem. Biol.*, 3:981-991 (1996))

There is, therefore, a need for an improved intein-mediated protein ligation method which overcomes the noted limitations of current IPL methods and which eliminates the need for use of proteases to generate an N-terminal cysteine residue. Such an improved IPL method would have widespread applicability for the ligation of expressed proteins, for example, labeling of extensive portions of a protein for, among other things, NMR analysis.

SUMMARY OF THE INVENTION

In accordance with the present invention, there is provided a method for the ligation of expressed proteins utilizing one or more inteins which display cleavage at their N- and/or C-termini. In accordance with the present invention, such inteins may occur

either naturally or may be modified to cleave at their N- and/or C-termini. Inteins displaying N- and/or C-terminal cleavage enable the facile isolation of a protein having a C-terminal thioester and a protein having an N-terminal amino acid residue such as cysteine, respectively, for use in the fusion of one or more expressed proteins. Alternatively, the method may be used to generate a single protein having both a C-terminal thioester and a specified N-terminal amino acid residue, such as cysteine, for the creation of cyclic or polymerized proteins. These methods involve the steps of generating at least one C-terminal thioester-tagged first target protein, generating at least one second target protein having a specified N-terminal amino acid residue, for example cysteine, and ligating these proteins. This method may be used where a single protein is expressed, where, for example, the C-terminal thioester end of the protein is fused to the N-terminal end of the same protein. The method may further include chitin-resin purification steps.

In one preferred embodiment the intein from the RIR1 *Methanobacterium thermoautotrophicum* is modified to cleave at either the C-terminus or N-terminus. The modified intein allows for the release of a bacterially expressed protein during a one-column purification, thus eliminating the need proteases entirely. DNA encoding these modified inteins and plasmids containing these modified inteins are also provided by the instant invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram depicting both the N-terminal and C-terminal cleavage reactions which comprise intein-mediated protein ligation. The modified *Mth* RIR1 intein was used to purify both MBP with a C-terminal thioester and T4 DNA ligase with an N-terminal cysteine. The *Mth* RIR1 intein for N-terminal cleavage, intein(N), carried the $P^{-1}G/N^{134}A$ double mutation. The full length fusion protein consisting of MBP-intein(N)-CBD was separated from cell extract by binding the CBD portion of the fusion protein to a chitin resin. Overnight incubation in the presence of 100 mM 2-mercaptoethanesulfonic acid (MESNA) induced cleavage of the peptide bond prior to the N-terminus of the intein and created a thioester on the C-terminus of MBP. The C-terminal cleavage vector, intein(C), had the $P^{-1}G/C^{1}A$ double mutation. The precursor CBD-intein(C)-T4 DNA ligase was isolated from induced *E. coli* cell extract by binding to a chitin resin as described for N-terminal cleavage. Fission of the peptide bond following the C-terminal residue of the intein at a preferred temperature and pH resulted in the production of T4 DNA ligase with an N-terminal cysteine. Ligation occurred when the proteins containing the complementary reactive groups were mixed and concentrated, resulting in a native peptide bond between the two reacting species.

Figure 2A is a gel depicting the purification of a C-terminal thioester-tagged maltose binding protein (MBP) via a thiol-inducible *Mth* RIR1 intein construct pMRB10G (containing the

modified intein, R(N), with $P^{-1}G/N^{134}A$ mutation) and the purification of T4 DNA ligase having an N-terminal cysteine using the vector pBRL-A (containing the modified intein, R(C), with $P^{-1}G/C^{1}A$ mutation). Lanes 1-3, purification of maltose binding protein (MBP) (M, 43 kDa) with a C-terminal thioester. Lane 1. ER2566 cells transformed with plasmid pMRB10G following Isopropyl ß-D-thiogalactopyranoside (IPTG) induction. Lane 2. Cell extract after passage over a chitin resin. Note that the fusion protein, M-R(N)-B, binds to the resin, where B is the chitin binding domain. Lane 3. Fraction 3 of the elution from the chitin resin following overnight incubation at 4°C in the presence of 100 mM MESNA. Lanes 4-6, purification of T4 DNA ligase (L, 56 kDa) with an N-terminal cysteine. Lane 4. IPTG induced ER2566 cells containing plasmid pBRL-A. Lane 5. Cell extract after application to a chitin resin. B-R(C)-L, the fusion protein, binds to the resin. Lane 6. Elution of T4 DNA ligase with an N-terminal cysteine after overnight incubation at room temperature in pH 7 buffer

Figure 2B is a gel depicting ligation of T4 DNA ligase having an N-terminal cysteine to a C-terminal thioester tagged MBP. Lane 1. Thioester-tagged MBP. Lane 2. T4 DNA ligase with an N-terminal cysteine. Lane 3. Ligation reaction of MBP (0.8 mM) with T4 DNA ligase (0.8 mM), generating M-L, after overnight incubation at 4°C.

Figure 3 is a gel depicting the effect of induction temperature on the cleaving and/or splicing activity of the *Mth*

RIR1 intein or *Mth* RIR1 intein mutants. The *Mth* RIR1 intein or mutants thereof, with 5 native N- and C-terminal extein residues were induced at either 15°C or 37°C. The intein was expressed as a fusion protein (M-R-B, 63 kDa) consisting of N-terminal maltose binding protein (M, 43 kDa), the *Mth* RIR1 intein (R, 15 kDa) and at its C-terminus was the chitin binding domain (B, 5 kDa). Lanes 1 and 2. M-R-B with the unmodified *Mth* RIR1 intein. Note the small amount of spliced product (M-B, 48 kDa). Lanes 3 and 4. *Mth* intein with Pro^{-1} replaced with Ala, M-R-B($P^{-1}A$). Both spliced product (M-B) and N-terminal cleavage product (M) are visible. Lanes 5 and 6. Replacement of Pro^{-1} with Gly (M-R-B($P^{-1}G$)) showed some splicing as well as N- and C-terminal cleavage, M and M-R, respectively. Lanes 7 and 8. The Pro^{-1} to Gly and Cys^1 to Ser double mutant, M-R-B($P^{-1}G/C^1S$), displayed induction temperature dependent C-terminal cleavage (M-R) activity. Lanes 9 and 10. The M-R-B($P^{-1}G/N^{134}A$) mutant possessed only N-terminal cleavage activity producing M. The *Mth* intein or *Mth* intein -CBD fusion is not visible in this Figure.

Figure 4 is a nucleotide sequence (SEQ ID NO:23) comparison of wild type *Mth* RIR1 intein and synthetic *Mth* RIR1 intein indicating the location of 61 silent base mutations designed to increase expression in *E. coli*. DNA alignment of the wild type *Mth* RIR1 intein (top strand) and the synthetic *Mth* RIR1 intein (bottom strand). To increase expression levels in *E. coli*, 61 silent base changes were made in 49 seperate codons when

creating the synthetic gene. The first and last codons of the wild type *Mth* RIR1 intein are shown in bold.

DETAILED DESCRIPTION

The present invention provides a solution to the limitations of current intein-mediated ligation methods by eliminating the need for a synthetic peptide as a ligation partner, and providing a method which is suitable for the fusion one or more expressed proteins.

In general, any intein displaying N- and/or C-terminal cleavage at its splice junctions can be used to generate a defined N-terminus, such as cysteine as well as a C-terminal thioester for use in the fusion of expressed proteins. Inteins which may be used in practicing the present invention include those described in Perler, et al., *Nucleic Acids Res.*, 27(1):346-347 (1999).

In accordance with one preferred embodiment, an intein found in the ribonucleoside diphosphate reductase gene of *Methanobacterium thermoautotrophicum* (the *Mth* RIR1 intein) was modified for the facile isolation of a protein with an N-terminal cysteine for use in the *in vitro* fusion of two bacterially-expressed proteins. The 134-amino acid *Mth* RIR1 intein is the smallest of the known mini-inteins, and may be close to the minimum amino acid sequence needed to promote splicing (Smith et.al., *J. Bacteriol.* 179: 7135-7155 (1997)).

The *Mth* RIR1 intein has a proline residue on the N-terminal side of the first amino acid of the intein. This residue was previously shown to inhibit splicing in the Sce VMA intein (Chong et al., *J. Biol. Chem.* 273:10567-10577 (1998)). The intein was found to splice poorly in *E. coli* when this naturally occurring proline is present. Splicing proficiency increases when this proline is replaced with an alanine residue. Constructs that display efficient N- and C-terminal cleavage are created by replacing either the C-terminal asparagine or N-terminal cysteine of the intein, respectively, with alanine.

These constructs allow for the formation of an intein-generated C-terminal thioester on a first target protein and an intein-generated N-terminal cysteine on a second target protein. These complementary reactive groups may then be ligated via native chemical ligation to produce a peptide bond (Evans et al *supra* (1998), Muir et al *supra* (1998)). Alternatively, a single protein containing both reactive groups may be generated for the creation of cyclic or polymerized proteins. Likewise, more than one first or second target proteins may be generated via use of multiple mutant inteins.

As used herein, the terms fusion and ligation are used interchangeably. Also as used herein, protein shall mean any protein, fragment of any protein, or peptide capable of ligation according to the methods of the instant invention. Further, as used herein, target protein shall mean any protein the ligation of

which, according to the methods of the instant invention, is desired.

The general method of intein-mediated protein ligation in accordance with the present invention is as follows:

(1) An intein of interest is isolated and cloned into an appropriate expression vector(s) such as bacterial, plant, insect, yeast and mammalian cells.

(2) The intein is engineered for N- and/or C-terminal cleavage unless the wild type intein displays the desired cleavage activities. In a preferred embodiment, a modified intein with the desired cleavage properties can be generated by substituting one or more residues within and/or flanking the intein sequence. For example, a modified intein having N-terminal cleavage activity can be created by changing the last intein residue. Alternatively, a modified intein with C-terminal cleavage activity can be created by changing the first intein residue.

(3) The intein with N- and/or C-terminal cleavage activity is fused with an affinity tag to allow purification away from other endogenous proteins.

(4) The intein or inteins, either wild type or modified, that display N-terminal and/or C-terminal cleavage, or both, are fused to the desired target protein coding region or regions upstream and/or downstream of the intein.

(5) An intein that cleaves at its N-terminus in a thiol reagent dependent manner is used to isolate a protein with a C-terminal thioester. This cleavage and isolation is, for example, carried out as previously described for the *Sce* VMA and *Mxe* GyrA inteins (Chong et al., *Gene* 192(2):271-281 (1997); Evans et al., *Protein Sci.* 7:2256-2264 (1998)). As discussed previously, multiple C-terminal thioester-tagged proteins may be generated at this step .

(6) A target protein having a specified N-terminus is generated by cleavage of a construct containing an intein that cleaves at its C-terminus. The specified N-terminal residue may be any of the amino acids, but preferably cysteine. As discussed previously, this step may alternately generate a specified N-terminal on the same protein containing a C-terminal thioester, to yield a single protein containing both reactive groups. Alternatively, multiple proteins having the specified N-terminus may be generated at this step.

(7) Thioester-tagged target protein and target protein having a specified N-termini are fused via intein-mediated protein ligation (IPL) (see Figure 2B). In a preferred embodiment, the N-terminus is cysteine. Alternatively, a single protein containing both a C-terminal thioester and a specified N-terminus, such as a cysteine, may undergo intramolecular ligation to yield a cyclic product and/or intermolecular ligation to yield polymerized proteins.

The methodology described by the instant invention significantly expands the utility of current IPL methods to enable the labeling of extensive portions of a protein for NMR analysis and the isolation of a greater variety of cytotoxic proteins. In addition, this advance opens the possibility of labeling the central portion of a protein by ligating three or more fragments.

The use of an intein or inteins with N-terminal and C-terminal cleavage activity provides the potential to create a defined N-terminus, such as a cysteine, and a C-terminal thioester on a single protein. The intramolecular ligation of the resulting protein generates a circular protein, whereas the intermolecular ligation of several of these proteins generates a protein polymer.

Cleavage at the N- and/or the C-terminus of an intein can be brought about by introducing changes to the intein and/or its extein sequences. Also, naturally occuring inteins may display these properties and require no manipulation. Cleavage at the N- and/or C-terminus of an intein can occur uncontrollably or induced using nucleophilc compounds, such as thiol reagents, temperature, pH, salt, chaotropic agents, or any combination of the aforementioned conditions and/or reagents.

The Examples presented below are only intended as specific preferred embodiments of the present invention and are not intended to limit the scope of the invention except as

provided in the claims herein. The present invention encompasses modifications and variations of the methods taught herein which would be obvious to one of ordinary skill in the art.

The references cited above and below are herein incorporated by reference.

EXAMPLE I

Creation of the *Mth* RIR1 synthetic gene

The gene encoding the *Mth* RIR1 intein along with 5 native N- and C-extein residues (Smith et al. *supra* (1997)) was constructed using 10 oligonucleotides (New England Biolabs, Beverly, MA) comprising both strands of the gene, as follows:

1) 5'-TCGAGGCAACCAACCCCTGCGTATCCGGTGACACCATTGT AATGACTAGTGGCGGTCCGCGCACTGTGGCTGAACTGGAG GGCAAACCGTTCACCGCAC-3' (SEQ ID NO:1)

2) 5'-CCGGTTGGCTGCTCGCCACAGTTGTGTACAATGAAGCCAT TAGCAGTGAATGCGCTAGCACCGTAAACAGTAGCGTCATA AACATCCTGGCGG-3' (SEQ ID NO:2)

3) 5'-pTGATTCGCGGCTCTGGCTACCCATGCCCCTCAGGTTTCTT CCGCACCTGTGAACGTGACGTATATGATCTGCGTACACGT GAGGGTCATTGCTTACGTTT-3' (SRQ ID NO:3)

4) 5'-pGACCCATGATCACCGTGTTCTGGTGATGGATGGTGGCCTG GAATGGCGTGCCGCGGGTGAACTGGAACGCGGCGACCGCC TGGTGATGGATGATGCAGCT-3' (SEQ ID NO:4)

5) 5'-pGGCGAGTTTCCGGCACTGGCAACCTTCCGTGGCCTGCGTG GCGCTGGCCGCCAGGATGTTTATGACGCTACTGTTTACGG TGCTAGC-3' (SEQ ID NO:5)

6) 5'-pGCATTCACTGCTAATGGCTTCATTGTACACAACTGTGGCG AGCAGCCAA-3' (SEQ ID NO:6)

7) 5'-pCCAGCGCCACGCAGGCCACGGAAGGTTGCCAGTGCCGGAA ACTCGCCAGCTGCATCATCCATCACCAGGCGGTCGCCGCG TTCCAGTTCACCCGCGGCAC-3' (SEQ ID NO:7)

8) 5'-pGCCATTCCAGGCCACCATCCATCACCAGAACACGGTGATC ATGGGTCAAACGTAAGCAATGACCCTCACGTGTACGCAGA TCATATACGT-3' (SEQ ID NO:8)

9) 5'-pCACGTTCACAGGTGCGGAAGAAACCTGAGGGGCATGGGTA GCCAGAGCCGCGAATCAGTGCGGTGAACGGTTTGCCCTCC AGTTCAGCCACAGTGCG-3' (SEQ ID NO:9)

10) 5'-pCGGACCGCCACTAGTCATTACAATGGTGTCACCGGATACG CAGGGGTTGGTTGCC-3' (SEQ ID NO:10)

To ensure maximal *E. coli* expression, the coding region of the synthetic *Mth* RIR1 intein incorporates 61 silent base mutations in 49 of the 134 codons (see Figure 4) in the wildtype *Mth* RIR1 intein gene (GenBank AE000845). The oligonucleotides were annealed by mixing at equimolar ratios (400 nM) in a ligation buffer (50 mM Tris-HCl, pH 7.5 containing 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, and 25 μg BSA) followed by heating to 95°C. After cooling to room temperature, the annealed and ligated oligonucleotides were inserted into the *Xho*I

and *Age*I sites of pMYB5 (NEB), replacing the *Sce* VMA intein and creating the plasmid pMRB8P.

Engineering the *Mth* RIR1 intein for N- and C-terminal cleavage

The unique *Xho*I and *Spe*I sites flanking the N-terminal splice junction and the unique *BsrG*I and *Age*I sites flanking the C-terminal splice junction allowed substitution of amino acid residues by linker replacement. The proline residue, Pro^{-1}, preceding the intein in pMRB8P was substituted with alanine or glycine to yield pMRB8A and pMRB8G1, respectively. Substitution of Pro^{-1}-Cys^{1} with Gly-Ser or Gly-Ala yielded pMRB9GS and pMRB9GA, respectively. Replacing Asn^{134} with Ala in pMRB8G1 resulted in pMRB10G. The following linkers were used for substitution of the native amino acids at the splice junctions (each linker was formed by annealing two synthetic oligonucleotides as described above):

$P^{-1}A$ linker: 5'-TCGAGGCAACCAACGCATGCGTATCCGGT GACACCATTGTAATGA-3' (SEQ ID NO:11)

and 5'-CTAGTCATTACAATGGTGTCACCGGATAC GCATGCGTTGGTTGCC-3' (SEQ ID NO:12)

$P^{-1}G$ linker: 5'-TCGAGGGCTGCGTATCCGGTGACACCATT GTAATGA–3 (SEQ ID NO:13)'

and 5'-CTAGTCATTACAATGGTGTCACCGGATAC GCAGCCC-3' (SEQ ID NO:14)

P^{-1}G/C^{1}S linker: 5'-TCGAGGGCATCGAGGCAACCAACGGATC CGTATCCGGTGACACCATTGTAATGA-3' (SEQ ID NO:15)

and 5'-CTAGTCATTACAATGGTGTCACCGGATAC GGATCCGTTGGTTGCCTCGATGCCC-3' (SEQ ID NO:16)

P^{-1}G/C^{1}A linker: 5'-TCGAGGGCATCGAGGCAACCAACGGCGCC GTATCCGGTGACACCATTGTAATGA -3' (SEQ ID NO:17)

and 5'-CTAGTCATTACAATGGTGTCACCGGATAC GGCGCCGTTGGTTGCCTCGATGCCC-3' (SEQ ID NO:18)

N^{134}A linker: 5'-GTACACGCATGCGGCGAGCAGCCCGG GA-3' (SEQ ID NO:19)

and 5'-CCGGTCCCGGGCTGCTCGCCGCATGC GT-3' (SEQ ID NO:20)

pBRL-A was constructed by substituting the *Escherichia coli* maltose binding protein (MBP) and the *Bacillus circulans* chitin binding domain (CBD) coding regions in pMRB9GA with the CBD and the T4 DNA ligase coding regions, respectively, subcloned from the pBYT4 plasmid.

EXAMPLE II

Generating a thioester-tagged protein:

The pMRB10G construct from Example I contains the *Mth* RIR1 intein engineered to undergo thiol reagent induced cleavage at the N-terminal splice junction (Figure 1, N-terminal cleavage) and was used to isolate proteins with a C-terminal thioester as described previously for the *Sce* VMA and *Mxe* GyrA inteins (Chong et al. *supra* 1997); Evans et al., supra (1998)). Briefly, ER2566 cells (Evans et.al. (1998)) containing the appropriate plasmid were grown at 37°C in LB broth containing 100 μg/mL ampicillin to an OD_{600} of 0.5-0.6 followed by induction with IPTG (0.5 mM). Induction was either overnight at 15°C or for 3 hours at 30°C.

The cells were pelleted by centrifugation at 3,000xg for 30 minutes followed by resuspension in buffer A (20 mM Tris-HCl, pH 7.5 containing 500 mM NaCl). The cell contents were released by sonication. Cell debris was removed by centrifugation at 23,000xg for 30 minutes and the supernatant was applied to a column packed with chitin resin (10 mL bed volume) equilibrated in buffer A. Unbound protein was washed from the column with 10 column volumes of buffer A.

Thiol reagent-induced cleavage was initiated by rapidly equilibrating the chitin resin in buffer B (20 mM Tris-HCl, pH 8 containing 500 mM NaCl and 100 mM 2-mercaptoethane-sulfonic acid (MESNA)). The cleavage reaction, which simultaneously

generates a C-terminal thioester on the target protein, proceeded overnight at 4°C after which the protein was eluted from the column. The use of the pMRB10G construct resulted in the isolation of MBP with a C-terminal thioester (Figure2A).

Isolating proteins with an N-terminal cysteine

The pBRL-A construct from Example I contains an *Mth* RIR1 intein engineered to undergo controllable cleavage at its C-terminus, and was used to purify proteins with an N-terminal cysteine (Figure 1, C-terminal cleavage). The expression and purification protocol was performed as described in Example II, except with buffer A replaced by buffer C (20 mM Tris-HCl, pH 8.5 containing 500 mM NaCl) and buffer B replaced by buffer D (20 mM Tris-HCl, pH 7.0 containing 500 mM NaCl). Also, following equilibration of the column in buffer D the cleavage reaction proceeded overnight at room temperature.

The expression of plasmid pBRL-A resulted in the purification of 4-6 mg/L cell culture of T4 DNA ligase possessing an N-terminal cysteine (Figure 2A). Protein concentrations were determined using the Bio-Rad protein assay (Bio-Rad Laboratories, Inc., Hercules, CA).

EXAMPLE III

Protein-protein ligation using Intein-mediated Protein Ligation

Intein-mediated protein ligation (IPL) was used to fuse two proteins (Figure 2B). Freshly isolated thioester-tagged protein from Example II was mixed with freshly isolated protein containing an N-terminal cysteine residue from Example II, with typical starting concentrations of 1-200 μM. The solution was concentrated with a Centriprep 3 or Centriprep 30 apparatus (Millipore Corporation, Bedford, MA) then with a Centricon 3 or Centricon 10 apparatus to a final concentration of 0.15-1.2 mM for each protein.

Ligation reactions proceeded overnight at 4°C and were visualized using SDS-PAGE with 12% Tris-glycine gels (Novex Experimental Technology, San Diego, CA) stained with Coomassie Brilliant Blue. Typical ligation efficiencies ranged from 20-60%.

Confirmation of ligation in IPL reactions

A Factor Xa site in MBP that exists 5 amino acids N-terminal from the site of fusion (Maina et al, *supra* (1988)) allowed amino acid sequencing through the ligation junction. The sequence obtained was NH_2-TLEGCGEQPTGXLK-COOH (SEQ ID NO:21) which matched the last 4 residues of MBP (TLEG) followed by a linker sequence (CGEQPTG (SEQ ID NO:22)) and the start of T4 DNA ligase (ILK). During amino acid sequencing, the cycle expected to yield an isoleucine did not have a strong

enough signal to assign it to a specific residue, so it was represented as an X. The cysteine was identified as the acrylamide alkylation product.

The Factor Xa proteolysis was performed on 2 mg of ligation reaction involving MBP and T4 DNA ligase. This reaction mixture was bound to 3 mL of amylose resin (New England Biolabs, Inc., Beverly, MA) equilibrated in buffer A (see Example II). Unreacted T4 DNA ligase was rinsed from the column with 10 column volumes of buffer A. Unligated MBP and the MBP-T4 DNA ligase fusion protein were eluted from the amylose resin using buffer E (20 mM Tris-HCl, pH 7.5 containing 500 mM NaCl and 10 mM maltose). Overnight incubation of the eluted protein with a 200:1 protein:bovine Factor Xa (NEB) ratio (w/w) at 4°C resulted in the proteolysis of the fusion protein and regeneration of a band on SDS-PAGE gels that ran at a molecular weight similar to T4 DNA ligase. N-terminal amino acid sequencing of the proteolyzed fusion protein was performed on a Procise 494 protein sequencer (PE Applied Biosystems, Foster City, CA).

Temperature sensitivity of the *Mth* RIR1 intein

The cleavage and/or splicing activity of the *Mth* RIR1 intein was more proficient when protein synthesis was induced at 15°C than when the induction temperature was raised to 37°C (Figure 3). The effect temperature has on the *Mth* RIR1 represents a way to control the activity of this intein for use in controlled splicing or cleavage reactions. Replacement of Pro^{-1} with a Gly

and Cys^1 with a Ser resulted in a double mutant, the pMRB9GS construct, which showed only *in vivo* C-terminal cleavage activity when protein synthesis was induced at 15°C but not at 37°C. Another double mutant, the pMRB9GA construct, displayed slow cleavage, even at 15°C, which allowed the accumulation of substantial amounts of the precursor protein and showed potential for use as a C-terminal cleavage construct for protein purification.

WHAT IS CLAIMED IS:

1. A method for fusion of expressed proteins, said method comprising the steps of:

(a) generating at least one C-terminal thioester-tagged first target protein;

(b) generating at least one second target protein having a specified N-terminal; and

(c) ligating said first and said second target proteins.

2. The method of claim 1, wherein said first target protein of step (a) is generated from a first plasmid comprising at least one first intein having N-terminal cleavage activity and said second target protein of step (b) is generated from a second plasmid comprising at least one second intein having C-terminal cleavage activity.

3. The method of claim 2, wherein said first intein comprises a first modified *Mth* RIR1 intein and wherein said second modified intein comprises a second modified *Mth* RIR1 intein.

4. The method of claim 3, wherein said first modified *Mth* RIR1 intein is selected from the group consisting of a Pro^{-1} to Ala mutant intein, a Pro^{-1} to Gly mutant intein, and a Pro^{-1} - Asn^{134} to Gly-Ala mutant intein, and wherein said second modified *Mth* RIR1 intein is selected from the group

consisting of a Pro^{-1} - Cys^{1} to Gly-Ser mutant intein and a Pro^{-1} - Cys^{1} to Gly-Ala mutant intein.

5. The method of claim 3, wherein said first plasmid is selected from the group consisting of pMRB8A, pMRB8G1 and pMRB10G, and wherein said second plasmid is selected from the group consisting of pMRB9GS, pMRB9GA and pBRL-A.

6. The method of claim 3, wherein said first target protein of step (a) is generated by thiol reagent-induced cleavage of said first modified *Mth* RIR1 intein and said second target protein of step (b) is generated by temperature and/or pH induced cleavage of said second modified *Mth* RIR1 intein.

7. The method of claim 2, wherein said specified N-terminal of step (b) comprises cysteine.

8. A method for fusion of expressed proteins, said method comprising the steps of:

(a) constructing a first plasmid comprising at least one first target protein and at least one first modified intein, wherein said first modified intein is capable of thiol reagent-induced cleavage to produce a thioester at the C-terminal of said first target protein;

(b) constructing a second plasmid comprising at least one second target protein and at least one second intein having C-terminal cleavage activity, wherein said second intein is capable of cleavage to produce a said second target protein having a specified N-terminal;

(c) generating at least one C-terminal thioester-tagged first target protein from said first plasmid of step (a);

(d) generating at least one second target protein having a specified N-terminal from said second plasmid of step (b); and

(e) ligating said first target protein of step (c) with said second target protein of step (d).

9. The method of claim 8, wherein step (c) further comprises purifying said C-terminal thioester-tagged first protein and step (d) further comprises purifying said second target protein having a specified N-terminal.

10. The method of claim 9, wherein said purifications of step (c) and step (d) comprise purification on a chitin resin column.

11. The method of claim 8, wherein said first intein of step (a) comprises a first modified *Mth* RIR1 intein, and wherein said second intein of step (b) comprises a second modified *Mth* RIR1 intein.

12. The method of claim 11, wherein said first modified *Mth* RIR1 intein is selected from the group consisting of a Pro^{-1} to Ala mutant intein, a Pro^{-1} to Gly mutant intein, and a Pro^{-1} - Asn^{134} to Gly-Ala mutant intein, and wherein said second modified *Mth* RIR1 intein is selected from the group consisting of a Pro^{-1} - Cys^{1} to Gly-Ser mutant intein and a Pro^{-1} - Cys^{1} to Gly-Ala mutant intein.

13. The method of claim 12, wherein said first plasmid of step (a) is selected from the group consisting of pMRB8A, pMRB8G1 and pMRB10G, and wherein said second plasmid of step (b) is selected from the group consisting of pMRB9GS, pMRB9GA and pBRL-A.

14. The method of claim 8, wherein said specified N-terminal comprises cysteine.

15. A fusion protein produced by the method of any one of claims 1-14.

16. A method for cyclic fusion of an expressed protein, said method comprising the steps of:

(a) constructing a plasmid comprising at least one target protein, at least one first intein having N-terminal cleavage activity, and at least one second intein having C-terminal cleavage activity, wherein

said first intein is capable of thiol reagent-induced cleavage to produce a thioester at the C-terminal of said target protein and wherein said second intein is capable of cleavage to produce a specified amino acid at the N-terminal of said target protein;

(b) generating a C-terminal thioester-tagged target protein having a specified amino acid at its N-terminal from the plasmid of step (a); and

(c) ligating the N-terminus of said target protein to the C-terminus of said target protein to produce a cyclic protein.

17. A method for polymerization of an expressed protein, said method comprising the steps of:

(a) constructing a plasmid comprising at least one target protein, at least one first intein having N-terminal cleavage activity, and at least one second intein having C-terminal cleavage activity, wherein said first intein is capable of thiol reagent-induced cleavage to produce a thioester at the C-terminal of said target protein and wherein said second intein is capable of cleavage to produce a specified amino acid at the N-terminal of said target protein;

(b) generating a C-terminal thioester-tagged target protein having a specified amino acid at its N-terminal from the plasmid of step (a); and

(c) intermolecular ligation of said target proteins to yield a protein polymer.

18. The method of claim 16 or 17, wherein said first intein of step (a) comprises a first modified *Mth* RIR1 intein, and wherein said second intein of step (a) comprises a second modified *Mth* RIR1 intein.

19. The method of claim 18, wherein said first modified *Mth* RIR1 intein is selected from the group consisting of a Pro^{-1} to Ala mutant intein, a Pro^{-1} to Gly mutant intein, and a Pro^{-1} - Asn^{134} to Gly-Ala mutant intein, and wherein said second modified *Mth* RIR1 intein is selected from the group consisting of a Pro^{-1} - Cys^{1} to Gly-Ser mutant intein and a Pro^{-1} - Cys^{1} to Gly-Ala mutant intein.

20. The method of claim 16 or 17, wherein said specified amino acid comprises cysteine.

21. A cyclic protein produced by the method of any one of claim 16.

22. A modified intein comprising a mutant *Mth* RIR1 intein capable of thiol reagent-induced cleavage to produce a thioester at the C-terminal of an adjacent target protein.

23. The modified intein of claim 22, wherein said mutant *Mth* RIR1 intein is selected from the group consisting of a Pro^{-1}

to Ala mutant intein, a Pro^{-1} to Gly mutant intein, and a Pro^{-1} - Asn^{134} to Gly-Ala mutant intein.

24. A modified intein comprising a mutant intein capable of pH and temperature-induced cleavage to produce a specified residue at the N-terminal of an adjacent target protein.

25. The modified intein of claim 24, wherein said mutant intein comprises a mutant *Mth* R1R1 intein.

26. The modified intein of claim 25, wherein said specified residue is cysteine.

27. The modified intein of claim 25, wherein said mutant *Mth* R1R1 intein is selected from the group consisting of a Pro^{-1} - Cys^{1} to Gly-Ser mutant intein and a Pro^{-1} - Cys^{1} to Gly-Ala mutant intein.

28. A plasmid comprising at least one modified intein of any one of claims 22-27.

29. A plasmid comprising a modified *Mth* RIR1 intein, wherein said plasmid is selected from the group consisting of pMRB8P, pMRB8A, pMRB8G1, pMRB9GS, pMRB9GA, pMRB10G and pBRL-A.

30. A DNA segment encoding a modified *Mth* RIR1 intein, wherein said DNA segment is obtainable from a plasmid selected from the group consisting of pMRB8P, pMRB8A, pMRB8G1, pMRB9GS, pMRB9GA, pMRB10G and pBRL-A.

FIG. 1

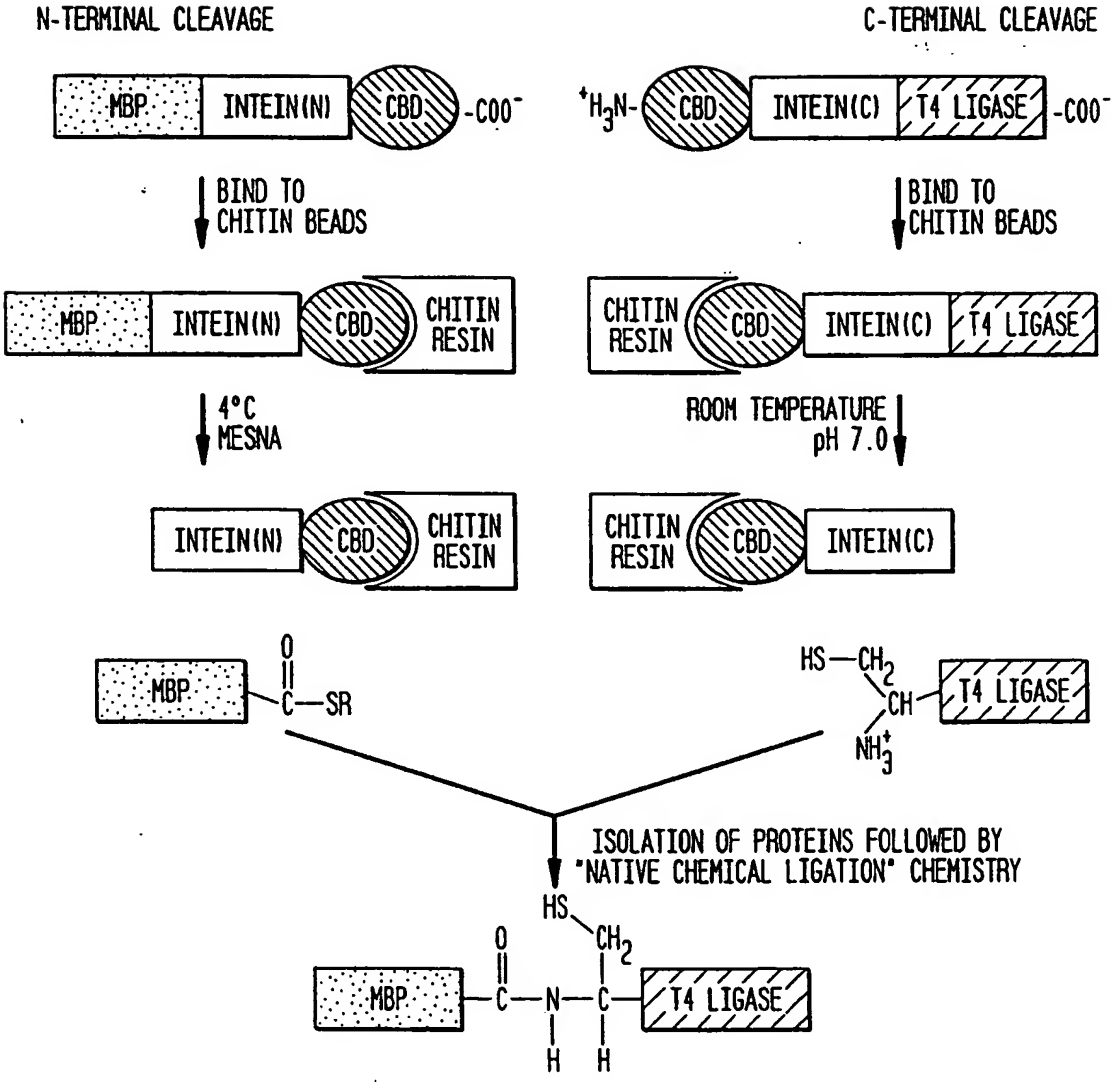

N-TERMINAL CLEAVAGE
C-TERMINAL CLEAVAGE
MBP
INTEIN(N)
CBD
-COO⁻
⁺H₃N-
INTEIN(C)
T4 LIGASE
BIND TO CHITIN BEADS
CHITIN RESIN
4°C MESNA
ROOM TEMPERATURE pH 7.0
ISOLATION OF PROTEINS FOLLOWED BY "NATIVE CHEMICAL LIGATION" CHEMISTRY

FIG. 2A

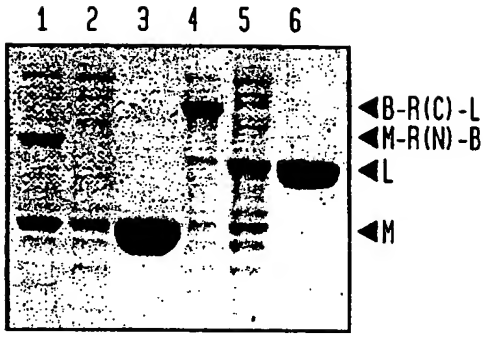


FIG. 2B

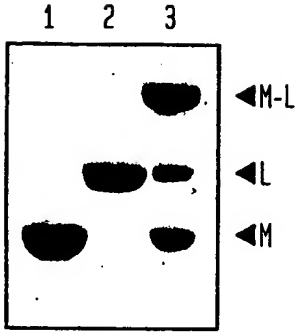


FIG. 3

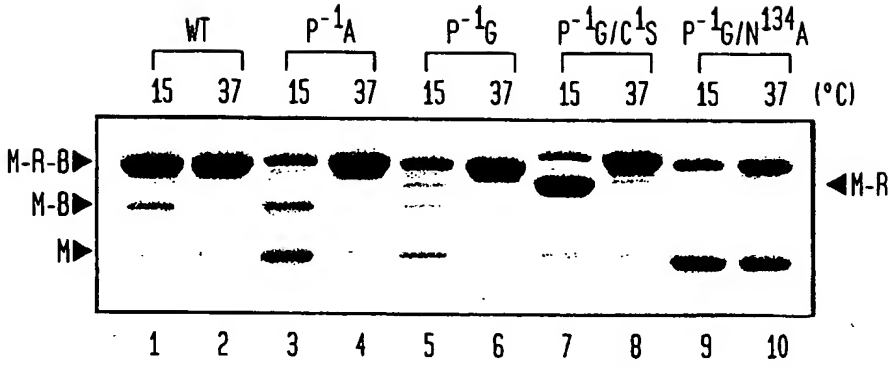
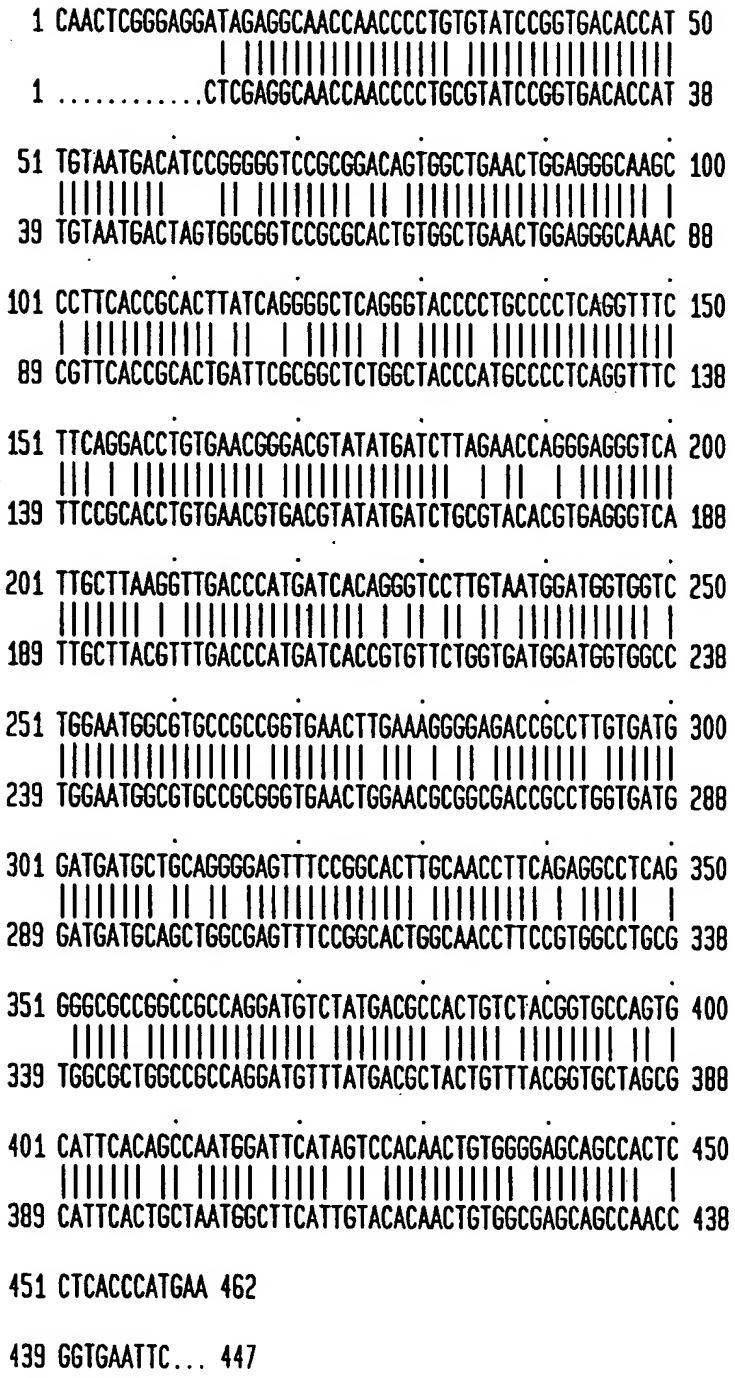

WT
P-1A
P-1G
P-1G/C1S
P-1G/N134A
15 37 15 37 15 37 15 37 15 37 (°C)
M-R-B▶
M-B▶
M▶
◀M-R
1 2 3 4 5 6 7 8 9 10

FIG. 4


1 CAACTCGGGAGGATAGAGGCAACCAACCCCTGTGTATCCGGTGACACCAT 50
1 .....CTCGAGGCAACCAACCCCTGCGTATCCGGTGACACCAT 38
51 TGTAATGACATCCGGGGGTCCGCGGACAGTGGCTGAACTGGAGGGCAAGC 100
39 TGTAATGACTAGTGGCGGTCCGCGCACTGTGGCTGAACTGGAGGGCAAAC 88
101 CCTTCACCGCACTTATCAGGGGCTCAGGGTACCCCTGCCCCTCAGGTTTC 150
89 CGTTCACCGCACTGATTCGCGGCTCTGGCTACCCATGCCCCTCAGGTTTC 138
151 TTCAGGACCTGTGAACGGGACGTATATGATCTTAGAACCAGGGAGGGTCA 200
139 TTCCGCACCTGTGAACGTGACGTATATGATCTGCGTACACGTGAGGGTCA 188
201 TTGCTTAAGGTTGACCCATGATCACAGGGTCCTTGTAATGGATGGTGGTC 250
189 TTGCTTACGTTTGACCCATGATCACCGTGTTCTGGTGATGGATGGTGGCC 238
251 TGGAATGGCGTGCCGCCGGTGAACTTGAAAGGGGAGACCGCCTTGTGATG 300
239 TGGAATGGCGTGCCGCGGGTGAACTGGAACGCGGCGACCGCCTGGTGATG 288
301 GATGATGCTGCAGGGGAGTTTCCGGCACTTGCAACCTTCAGAGGCCTCAG 350
289 GATGATGCAGCTGGCGAGTTTCCGGCACTGGCAACCTTCCGTGGCCTGCG 338
351 GGGCGCCGGCCGCCAGGATGTCTATGACGCCACTGTCTACGGTGCCAGTG 400
339 TGGCGCTGGCCGCCAGGATGTTTATGACGCTACTGTTTACGGTGCTAGCG 388
401 CATTCACAGCCAATGGATTCATAGTCCACAACTGTGGGGAGCAGCCACTC 450
389 CATTCACTGCTAATGGCTTCATTGTACACAACTGTGGCGAGCAGCCAACC 438
451 CTCACCCATGAA 462
439 GGTGAATTC... 447

SEQUENCE LISTING

<110> Evans, Thomas
Xu, Ming-Qun
NEW ENGLAND BIOLABS, INC.

<120> Intein-Mediated Protein Ligation Of Expressed Proteins

<130> NEB-154-PCT

<140>
<141>

<150> 09/249,543
<151> 1999-02-12

<160> 24

<170> PatentIn Ver. 2.0

<210> 1
<211> 99
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically
Synthesized From Methanobacterium
thermoautotrophicum.

<400> 1
tcgaggcaac caacccctgc gtatccggtg acaccattgt aatgactagt ggcggtccgc 60
gcactgtggc tgaactggag ggcaaaccgt tcaccgcac 99

<210> 2
<211> 93
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically
Synthesized From Methanobacterium
thermoautotrophicum.

<400> 2
ccggttggct gctcgccaca gttgtgtaca atgaagccat tagcagtgaa tgcgctagca 60
ccgtaaacag tagcgtcata aacatcctgg cgg 93

<210> 3
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically
Synthesized From Methanobacterium

thermoautotrophicum.

<400> 3
tgattcgcgg ctctggctac ccatgcccct caggtttctt ccgcacctgt gaacgtgacg 60
tatatgatct gcgtacacgt gagggtcatt gcttacgttt 100

<210> 4
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 4
gacccatgat caccgtgttc tggtgatgga tggtggcctg gaatggcgtg ccgcgggtga 60
actggaacgc ggcgaccgcc tggtgatgga tgatgcagct 100

<210> 5
<211> 87
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 5
ggcgagtttc cggcactggc aaccttccgt ggcctgcgtg gcgctggccg ccaggatgtt 60
tatgacgcta ctgtttacgg tgctagc 87

<210> 6
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 6
gcattcactg ctaatggctt cattgtacac aactgtggcg agcagccaa 49

<210> 7
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 7
ccagcgccac gcaggccacg gaaggttgcc agtgccggaa actcgccagc tgcatcatcc 60
atcaccaggc ggtcgccgcg ttccagttca cccgcggcac 100

<210> 8
<211> 90
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically
Synthesized From Methanobacterium
thermoautotrophicum.

<400> 8
gccattccag gccaccatcc atcaccagaa cacggtgatc atgggtcaaa cgtaagcaat 60
gaccctcacg tgtacgcaga tcatatacgt 90

<210> 9
<211> 97
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically
Synthesized From Methanobacterium
thermoautotrophicum.

<400> 9
cacgttcaca ggtgcggaag aaacctgagg ggcatgggta gccagagccg cgaatcagtg 60
cggtgaacgg tttgccctcc agttcagcca cagtgcg 97

<210> 10
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically
Synthesized From Methanobacterium
thermoautotrophicum.

<400> 10
cggaccgcca ctagtcatta caatggtgtc accggatacg caggggttgg ttgcc 55

<210> 11
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically
Synthesized From Methanobacterium
thermoautotrophicum.

<400> 11
tcgaggcaac caacgcatgc gtatccggtg acaccattgt aatga 45

<210> 12
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 12
ctagtcatta caatggtgtc accggatacg catgcgttgg ttgcc 45

<210> 13
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 13
tcgagggctg cgtatccggt gacaccattg taatga 36

<210> 14
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 14
ctagtcatta caatggtgtc accggatacg cagccc 36

<210> 15
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 15
tcgagggcat cgaggcaacc aacggatccg tatccggtga caccattgta atga 54

<210> 16

<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 16
ctagtcatta caatggtgtc accggatacg gatccgttgg ttgcctcgat gccc 54

<210> 17
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 17
tcgagggcat cgaggcaacc aacggcgccg tatccggtga caccattgta atga 54

<210> 18
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 18
ctagtcatta caatggtgtc accggatacg gcgccgttgg ttgcctcgat gccc 54

<210> 19
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 19
gtacacgcat gcggcgagca gcccggga 28

<210> 20
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 20
ccggtcccgg gctgctcgcc gcatgcgt 28

<210> 21
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<220>
<223> At position 12, "Xaa" = any amino acid

<400> 21
Thr Leu Glu Gly Cys Gly Glu Gln Pro Thr Gly Xaa Leu Lys
1 5 10

<210> 22
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 22
Cys Gly Glu Gln Pro Thr Gly
1 5

<210> 23
<211> 462
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically Synthesized From Methanobacterium thermoautotrophicum.

<400> 23
caactcggga ggatagaggc aaccaacccc tgtgtatccg gtgacaccat tgtaatgaca 60
tccgggggtc cgcggacagt ggctgaactg gagggcaagc ccttcaccgc acttatcagg 120
ggctcagggt acccctgccc ctcaggtttc ttcaggacct gtgaacggga cgtatatgat 180
cttagaacca gggagggtca ttgcttaagg ttgacccatg atcacagggt ccttgtaatg 240

gatggtggtc tggaatggcg tgccgccggt gaacttgaaa ggggagaccg ccttgtgatg 300
gatgatgctg caggggagtt tccggcactt gcaaccttca gaggcctcag gggcgccggc 360
cgccaggatg tctatgacgc cactgtctac ggtgccagtg cattcacagc caatggattc 420
atagtccaca actgtgggga gcagccactc ctcacccatg aa 462

<210> 24
<211> 447
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Chemically
Synthesized From Methanobacterium
thermoautotrophicum.

<400> 24
ctcgaggcaa ccaacccctg cgtatccggt gacaccattg taatgactag tggcggtccg 60
cgcactgtgg ctgaactgga gggcaaaccg ttcaccgcac tgattcgcgg ctctggctac 120
ccatgcccct caggtttctt ccgcacctgt gaacgtgacg tatatgatct gcgtacacgt 180
gagggtcatt gcttacgttt gacccatgat caccgtgttc tggtgatgga tggtggcctg 240
gaatggcgtg ccgcgggtga actggaacgc ggcgaccgcc tggtgatgga tgatgcagct 300
ggcgagtttc cggcactggc aaccttccgt ggcctgcgtg gcgctggccg ccaggatgtt 360
tatgacgcta ctgtttacgg tgctagcgca ttcactgcta atggcttcat tgtacacaac 420
tgtggcgagc agccaaccgg tgaattc 447

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/02764

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.7, 69.1, 252.3, 320.1, 471; 530/350, 402, 408, 412, 413; 536/23.4

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAS ONLINE: Chemical Abstracts

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	XU. R. et al. Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies. Proceedings of the National Academy of Sciences, U.S.A. January 1999. Vol. 96. No. 2. pages 388-393, especially Figure 1 and pages 388-392.	1 and 15 2-4,6-12,14, 16-28.
Y	US 5,834,247 A (COMB et al) 10 November 1998, Figures 28 and 37 and cols. 9-13, 36-65, and 73-89.	2-4,6-12,14, and 16-28.
Y	WU. H. et al. Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. Biochimica et Biophysica Acta. September 1998. Vol. 1387. Nos. 1-2. pages 422-432, especially Figures 2 and 4 and pages 423-432.	2-4,6-12,14, and 16-28.

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 12 JUNE 2000	Date of mailing of the international search report 19 JUL 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer WILLIAM W. MOORE Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/02764

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WU. H. et al. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. Proceeedings of the National Academy of Sciences, U.S.A. August 1998. Vol. 95. No. 15. pages 9226-9231, especially Figures 2 and 3 and pages 9226-9230.	2-4,6-12,14, and 16-28.
Y	SMITH. D.R. et al. Complete Genome Sequence of Methanobacterium thermoautotrophicum (delta)H: Functional Analysis and Comparative Genomics. Journal of Bacteriology. November 1997. Vol. 179. No. 22. pages 7135-7155, especially Figure 8 and pages 7135-7138 and 7152-7153.	3, 4, 6, 11, 12, 18, 19, 22, 23, 27, and 28.
Y	CHONG. S. et al. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. Nucleic Acids Research. November 1998. Vol. 26. No. 22. pages 5109-5115, especially Table 2 and pages 5111-5115.	4, 6, 12, 19, 23, 24, and 27.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/02764

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

C12N 15/62, 15/63, 15/70; C12P 21/00, 21/04; C07K 19/00, 1/113, 1/14, 1/22

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/69.7, 69.1, 252.3, 320.1, 471; 530/350, 402, 408, 412, 413; 536/23.4